STAT 1361: Statistical Learning and Data Science
Spring 2025

Time: M/W 12:00 – 1:15pm Room: 104 Lawrence Hall

Instructor: Prof. Lucas Mentch
Department of Statistics, University of Pittsburgh
Email: LKM31@pitt.edu Office: 1807 WW Posvar Hall Phone: 412-624-8730
Office Hours: Wed: 11:00am-12:00pm and by appt

Head TA: Ryan Cecil
Email: RMC144@pitt.edu Office: 1824 WW Posvar Hall
Office Hours: Th: 11:00-1:00pm & and by appt

TA/Grader: Alex Dukart
Email: ALD349@pitt.edu Office: WW Posvar Hall
Office Hours: By appt only

Course Delivery: The plan for this semester is to deliver lectures entirely in person on the days/times listed above. Students should be aware, however, that university policy may change throughout the semester and we will update the course as necessary to comply with any restrictions that are added. If at some point throughout the semester the instructor(s) are sick or otherwise unable to deliver in-person lectures, lectures may be given over zoom at the regularly-scheduled class time. In such an event, zoom links and recordings will be provided through the course webpage (canvas). Any and all changes to the status of the course will be posted on canvas throughout the semester.

Course description: This course is designed to provide a broad introduction to the field of data science and to expose students to many of the statistical tools most commonly used by modern data scientists. We will explore a wide variety of models and algorithms in a data-driven fashion. Topics will include modeling techniques ranging from classic statistical modeling (e.g. linear and logistic regression) to modern statistical learning (e.g. regularization and lasso) to fundamental machine learning (e.g. random forests and support vector machines). Particular attention will be given to the sorts of scientific questions that can be asked and answered within the different frameworks. Students will have the opportunity to utilize modern, interesting datasets to both provide data-driven analytical solutions and also to formally assess the uncertainty in making such determinations. The R language will be used extensively for statistical computing. Some prior knowledge or experience with R or related programming languages is helpful but not essential.

Textbooks: "An Introduction to Statistical Learning: with Applications in R." James, G., Witten, D.,Hastie, T., Tibshirani, R. Springer. 2^{nd} Edition. ISBN: 978-1071614174 (Required)

We will follow the above textbook very closely – note the new 2^{nd} edition. For more detail on some methods and R functionality, you may also be interested in referencing the following textbook, though it is not required:

"Modern Data Science with R." Baumer, B., Kaplan D., Horton J. CRC Press. ISBN: 978-1498724487

Prerequisite: STAT 1261 or equivalent (e.g. STAT 1221 and some programming experience).

Programming Help:

- While not formally necessary, students are generally expected to be comfortable programming in R prior to entering the course. Homework assignments will nearly always include extensive R programming, but this aspect will rarely if ever be discussed in lectures. R programming is *not* taught as part of the course. Students entering the course without significant experience in R should work diligently during the first several weeks of the course to get the basics down before homeworks are due.

  Students may, of course, ask the professor and TAs R-related questions relevant to the course material, but should look for other resources if more extensive help and/or troubleshooting coding-related issues is needed. One option is the Statistics Department Computing Lab ("Stat Lab") – more info can be found on the statistics department website. The Pitt Library also frequently offers an open office hour on R and RStudio; more info can be found here: `https://dojobo.github.io/r-lessons-learned/index.html`.

Course Requirements and Grading:

- Homework – 50%
- Quizzes – 30%
- Final Project – 20%

Class Participation:

- Students are expected to regularly attend lectures and actively participate in class discussions.

Homework Policy:

- Homework assignments will be uploaded to canvas and should be completed and submitted by the designated date/time. Specific submission instructions will be provided for each assignment. Approximately 7 homework assignments will be given throughout the semester. For select topics, supplemental labs and/or online tutorials will also be provided.

  Except in the case of emergencies or prior written approval, late homework will not be accepted.

- In questions that involve R code, all relevant code necessary to reproduce your results must be provided. Homeworks should be neatly done and organized in an easy-to-read fashion – you should not be submitting your first attempt at solutions. Part of your homework grade will be based on the presentation of solutions. The course instructor and TA reserve the right to refuse any homework that does not follow these guidelines.

Quiz Policy:

- Quizzes are designed to ensure a high-level understanding of the material covered in previous weeks. A total of approximately 5 quizzes will be given **at the start of class** throughout the semester and students will have approximately 20 minutes at the start of class to complete them. (Students who arrive late will not be given extra time.) Quizzes may occasionally be administered online through canvas if the course is operating remotely during a scheduled quiz time.
- Students will be given advance notice of quizzes.
- Quizzes are closed book and closed note. **No aids of any kind are permitted.**

Final Project:

- A primary component of this course (25% of your final grade) is an independent final data science project in which you will be tasked with using the knowledge acquired throughout the semester to complete a data analysis and report from start to finish. More information will be provided on the course webpage and in class discussions.

University Policies:

- *Academic Integrity*

  Students in this course will be expected to comply with the University of Pittsburgh's Policy on Academic Integrity. Any student suspected of violating this obligation for any reason during the semester will be required to participate in the procedural process, initiated at the instructor level, as outlined in the University Guidelines on Academic Integrity. This may include, but is not limited to, the confiscation of the examination of any individual suspected of violating University Policy. Furthermore, no student may bring any unauthorized materials to an exam, including dictionaries and programmable calculators.

- *Disability Services*

  If you have a disability for which you are or may be requesting an accommodation, you are encouraged to contact both your instructor and Disability Resources and Services (DRS), 140 William Pitt Union, (412) 648-7890, drsrecep@pitt.edu, (412) 228-5347 for P3 ASL users, as early as possible in the term. DRS will verify your disability and determine reasonable accommodations for this course.

- *Copyright Notice*

  Course materials may be protected by copyright. United States copyright law, 17 USC section 101, et seq., in addition to University policy and procedures, prohibit unauthorized duplication or retransmission of course materials. See Library of Congress Copyright Office and the University Copyright Policy.

- *E-mail Communication*

  Each student is issued a University e-mail address (username@pitt.edu) upon admittance. This e-mail address may be used by the University for official communication with students. Students are expected to read e-mail sent to this account on a regular basis. Failure to read and react to University communications in a timely manner does not absolve the student from knowing and complying with the content of the communications. The University provides an e-mail forwarding service that allows students to read their e-mail via other service providers (e.g., Hotmail, AOL, Yahoo). Students that

choose to forward their e-mail from their pitt.edu address to another address do so at their own risk. If e-mail is lost as a result of forwarding, it does not absolve the student from responding to official communications sent to their University e-mail address. To forward e-mail sent to your University account, go to http://accounts.pitt.edu, log into your account, click on Edit Forwarding Addresses, and follow the instructions on the page. Be sure to log out of your account when you have finished. (For the full E-mail Communication Policy, go to www.bc.pitt.edu/policies/policy/09/09-10-01.html.)